

REMARKS

Claims 1-20 have been cancelled, and new claims 21-28 have been added. Claims 21-28 are pending.

The specification has been amended to make the Title more descriptive of the claimed invention and to remove the "Field of the Invention" paragraph.

In the Office Action, the drawings were objected to for failing to show how text-to-speech conversion is achieved and how vocabulary domains are made and compared against input text, and also for failing to illustrate speech grammar. It is respectfully requested that these rejections be withdrawn for the following reasons.

With respect to the speech grammar, claims 6 and 16 which recited the speech grammar have been cancelled, and new claims 21-28 do not recite any speech grammar. Thus, the lack of speech grammar in the drawings is not seen to be pertinent to the invention as now claimed.

With respect to the conversion of text to speech, it is respectfully submitted that the drawings in combination with the text describe how text-to-speech conversion is performed, at least insofar as pertinent to the invention as set forth in claims 21-28. Figure 2 shows a set of components including a vocabulary domain based text-to-speech converter 210, IVR 180, and TTS server 220, and the text on pages 6-7 describe the functions performed by these elements when text is to be converted to speech. Additionally, Figure 3 and the accompanying text describe more specific functionality of the TTS server 220 in conjunction with limited vocabulary domain servers 340, which involves parsing input text into phrases, obtaining corresponding audio content from a cache 330 if it was previously cached, mapping the phrases to respective "clusters" or vocabulary domains that have been previously defined, providing each phrase to a limited vocabulary domain server 340 capable of converting the phrase to audio content, and transmitting the converted phrase back to a requesting client 300, which in the embodiment of Figure 2 is the IVR 180 which actually renders the audio content to the subscriber 200. The manner of conversion essentially

involves identifying audio files (such as .wav files, etc.) stored in an audio database 130 based on the parsed text, and then rendering the audio files to a user via an IVR 180. It is respectfully submitted that the existing drawings in combination with the accompanying text describe this process of text-to-speech conversion.

With respect to the definition and use of the vocabulary domains, it is respectfully submitted that the drawings in combination with the accompanying text describe these aspects of the invention as well. Figure 2 shows a domain definer 150 and its connections to other elements of the system including a TTS HTTP server 110, a service provider HTTP server 120, a recording studio 160 and audio database 130, and a user database 170. Paragraph 29 of the text describes that the domain definer 150 can group sentences that share one or more selected words into the same limited vocabulary domain, and gives the examples of "weather", "city-state information" and "customer information" as limited vocabulary domains that might be defined in this manner. Figure 3 and the accompanying text describe that a text parser 310 maps parsed phrases to these limited vocabulary domains, which involves for example determining whether each phrase includes a word or words that are part of the definition of a limited vocabulary domain, and a text distributor 320 enqueues the phrases on queues 350 associated with the respective domains for conversion by respective limited vocabulary domain servers 340.

It is further noted that new claims 21-28 do not include the step of converting the parsed phrases to audio content, but rather recite a second operation in which the parsed phrases are provided to a limited vocabulary domain server for conversion, and the audio components received from the limited vocabulary domain servers are used to generate the audio to the user. The actual conversion is not recited in the claim. Thus, to the extent that specific details of any conversion process used by the limited vocabulary domain servers are not shown in the application, such is not seen to be a basis for objecting to

the drawings or specification, because it is not pertinent to the invention as set forth in the claims.

Based on the foregoing, it is respectfully submitted that the drawings in conjunction with the text of the application show all the pertinent features of the invention as claimed. Accordingly, withdrawal of the objections to the drawings is respectfully requested.

In the Office Action, the "Field of the Invention" statement in the specification was objected to. This statement has been deleted.

In the Office Action, claims 2-3 and 12-13 were rejected under 35 U.S.C. § 112 2nd paragraph as being indefinite for the use of a "closeness metric". These specific claims have been cancelled, and therefore this rejection is no longer applicable to them. It is noted that new claims 21 and 25 also employ the term "closeness metric", but this term is not seen to render these claims indefinite. The Office Action incorrectly states that the only description of a closeness metric is in the Summary on page 2 of the application. However, the Summary indicates that the closeness metric is utilized to define the vocabulary domains, and the process of defining vocabulary domains is explained in some detail in paragraph 29 (page 5) of the application. In particular, it is described that each limited domain contains a cluster of similar vocabularies, such as sentences that share one or more selected words. Specific examples of "weather", "city-state information" and "customer information" are given for defining different limited vocabulary domains. It is respectfully submitted that this description makes it clear that the closeness metric is that which is used to define similar vocabularies, such as sharing one or more words, and that the general discussion and specific examples given in this section and elsewhere render this meaning of "closeness metric" sufficiently definite to satisfy 35 U.S.C. § 112, 2nd paragraph.

In the Office Action, claims 1-20 were rejected under 35 U.S.C. § 112, 1st paragraph, as containing subject matter not enabled by the specification. It is respectfully submitted that this rejection has been rendered moot by the

cancellation of these claims. To the extent that any of the specific objections in this rejection might be seen as applicable to new claims 21-28, such are addressed below.

With respect to the contention that the specification lacks specificity regarding how input text is converted into speech, this is seen as essentially the same as the objection to the drawings addressed above. It is believed that the above discussion regarding the teaching of the specification addresses this aspect of the rejection. As also noted above, paragraph 29 of the specification explains how the limited domains are different from each other and what they might contain.

With respect to the text distributor 320, it is not seen how the description is inadequate. As explained above, the limited vocabularies are defined by content in the text, such as specific words, and thus the distribution is simply a matter of sending each parsed phrase from the parser 310 to the appropriate server 340 based on its contents. If a phrase contains the word "weather", for example, then the phrase is sent to the server 340 that is responsible for the weather-related limited vocabulary.

The TTS clients 370 is described as being responsible for obtaining a phrase from a thread of the thread pool 360 and forwarding it to the respective limited domain server 340. The TTS clients 370 thus have limited roles of interfacing the TTS server 220 to separate limited domain servers 340. This is clearly distinguished from the client 300 that interfaces to one TTS server 220 whose functionality is not limited to a limited vocabulary domain. An example of a client 300 for the TTS server 220 is the IVR 180 shown in Figure 2, which is supplying general text for conversion and not aware of the division of labor that occurs within the TTS server, i.e., the parsing and distribution of text phrases to different limited domain vocabulary servers 340 based on their content.

Moreover, a client such as the IVR 180 plays a broader role in the system, being responsible for obtaining audio content for a web page, for example, and then creating the actual corresponding audio signals that are provided to the user. It

is hoped that this explanation of the differences between the TTS clients 370 and the client 300 is helpful.

With respect to the questions concerning details of the text-to-speech conversion, it has previously been explained that the process involves obtaining audio files corresponding to text phrases and then rendering the audio from the audio files to the user as audio signals. An example of such audio files are .wav files. Using such an approach, it is generally known that a collection of text will be rendered using a sequence of multiple audio files, effectively concatenating multiple short analog signals together. It is believed that this functionality is fairly described in or inferred from the specification. Additionally, as noted above, new claims 21-28 are not directed to the actual text-to-speech conversion process as performed within the limited domain servers 340, but rather to the overall process by which a user can obtain information from a text-based web site in audio form, which is performed by the entire collection of components illustrated in the Figures. Thus, any lack of specific details with respect to how audio components are generated from text phrases is not seen to render claims 21-28 unpatentable under 35 U.S.C. § 112, 1st paragraph.

In the Office Action, claims 1-20 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Saylor (US 6,707,889) and Oh (6,141,642). This rejection is moot due to the cancellation of claims 1-20. The pertinence of these references to new claims 21-28 is discussed below.

Claim 21 is a method of enabling a user to obtain information from a text-based web site in audio form. A first operation to prepare the text-based web site for delivery in audio form includes the following steps:

- (i) accessing content of a text-based web site to collect a vocabulary of textual information appearing therein;

- (ii) analyzing the collected vocabulary to determine a plurality of limited vocabulary domains into which the textual information of the web site can be grouped, the textual information of each limited vocabulary domain sharing a content-based closeness metric;

(iii) comparing the limited vocabulary domains with existing recorded audio content to determine whether additional audio content is necessary to deliver the web site in audio form, and if so then obtaining such additional audio content; and

(iv) storing formatting configuration information specifying how to deliver the text-based web site in audio format according to the limited vocabulary domains using the existing and additional audio content.

A second operation in claim 21 is performed upon a user's request for audio delivery of textual information from the text-based web site includes the following steps:

(i) obtaining the requested textual information from the text-based web site and parsing the textual information into phrases;

(ii) based on the stored formatting configuration information, mapping the parsed phrases to respective ones of the vocabulary domains and providing each parsed phrase to a corresponding limited vocabulary domain server capable of converting the parsed phrase to an audio component;

(iii) receiving audio components from the limited vocabulary domain servers, the audio component resulting from the conversion of the parsed phrases by the limited vocabulary domain servers; and

(iv) generating audio to the user based on the audio components received from the limited vocabulary domain servers.

Thus, claim 21 recites a method in which any rendering of text in audio form is preceded by an operation in which the text of the web site has been analyzed and grouped into limited vocabulary domains. This creation of limited vocabulary domains enhances the efficiency with which the text is converted to audio form.

Saylor is seen to teach a system and method in which voice pages (Vpages) are output to a user based on the input of a "voice code" (VCode). As

noted in the Office Action, Saylor does not provide any teaching with respect to limited vocabulary domains. Moreover, contrary to the assertion in the Office Action, Saylor is not seen to teach or suggest any selection of "an appropriate TTS engine" based on analyzing text of a web site. Col. 29 lines 22 et seq. of Saylor, which are referred to in the Office Action, describe that the "voice browser" 35 handles TTS processing through interaction with a TTS module, and determines the voice content that is desired based on translating voice input, not based on processing the text of the web site. Moreover, Saylor's mention of "an appropriate TTS engine" is not seen to disclose the claimed vocabulary domains, because nowhere in Saylor is there any mention of multiple TTS engines or any set of distinct specialized functions that are to be performed by multiple TTS engines. For all that is known, a TTS engine in Saylor may be "appropriate" because it is associated with a particular Vpage or Vcode, for example, and not because it is associated with a limited vocabulary domain. And in any event, any selection of a TTS engine in Saylor is based on voice input, and not based on analyzing text of a web site as set forth in claim 21.

Oh is seen to teach dividing input text into sub-texts according to language and using a plurality of text-to-speech engines, one for each language, for converting the divided sub-texts into audio wave data. The division of text in Oh is according to the identity of a language, not according to limited vocabulary domains that are defined by closeness metrics such as shared words. Moreover, in Oh there is no preceding analysis of the text in order to create limited vocabulary domains, which of course is unnecessary because the whole premise of Oh is that the text is already divided into English and Korean words/phrases. Thus, Oh fails to teach or suggest the elements of claim 21 that pertain to the creation of the limited vocabulary domains and the nature of the limited vocabulary domains (i.e., how they are defined).

It is respectfully submitted that claim 21 is not obvious in view of Saylor and Oh, because these references do not individually or collectively teach or suggest all the elements of claim 21. Neither of these references teaches or

suggests analyzing a vocabulary collected from a text-based web site to determine a plurality of limited vocabulary domains into which the textual information of the web site can be grouped, wherein the textual information of each limited vocabulary domain shares a content-based closeness metric. Additionally, neither of these references teaches or suggests mapping parsed text phrases to respective ones of the limited vocabulary domains and providing each parsed phrase to a corresponding limited vocabulary domain server capable of converting the parsed phrase to an audio component. In Saylor, there is only a vague suggestion of the use of multiple TTS engines based on voice input, and no description of mapping text phrases to multiple engines based on limited vocabulary domains that were determined in a preceding operation of analyzing the textual information of the web site. Oh teaches the use of multiple TTS engines solely to render different languages, not to render phrases from different vocabulary domains as defined in a preceding operation of analyzing textual information of the web site. Because these references fail to teach or suggest these features of claim 21, they are not seen to render claim 21 obvious under 35 U.S.C. § 103(a).

The remaining claims incorporate, either directly or indirectly, the above-discussed features of claim 21, and therefore are seen to be allowable in view of Saylor and Oh as well.

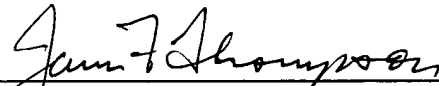
Based on the foregoing, it is believed that all objections and rejections have been addressed by this amendment and that this application is presently in condition for allowance. Favorable action is respectfully requested. If there should be any issues remaining after this amendment, the Examiner is urged to contact the undersigned attorney by telephone to resolve such issues.

If the U.S. Patent and Trademark Office deems a fee necessary, it may be charged to the account of the undersigned, Deposit Account No. 50-0901.

-15-

If the enclosed papers or fees are considered incomplete, the Patent Office is respectfully requested to contact the undersigned collect at (508) 366-9600, in Westborough, Massachusetts.

Respectfully submitted,



James F. Thompson, Esq.
Attorney for Applicant(s)
Registration No.: 36,699
CHAPIN & HUANG, L.L.C.
Westborough Office Park
1700 West Park Drive
Westborough, Massachusetts 01581
Telephone: (508) 366-9600
Facsimile: (508) 616-9805
Customer No.: 022468

Attorney Docket No.: NMS03-13

Dated: April 15, 2005